Rec'd PCT/PTO 19 JAN 2005
IB 03/03012
#2

Europäisches Patentamt
European Patent Office
Office européen des brevets

08 JUL 2003

REC'D 05 AUG 2003
WIPO PCT

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02291872.6

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

BEST AVAILABLE COPY

Anmeldung Nr:
Application no.: 02291872.6
Demande no:

Anmeldetag:
Date of filing: 24.07.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Method and encoder for processing a digital video signal.

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/26

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

FIELD OF THE INVENTION

The present invention relates to a method for processing a digital video signal, said digital video signal comprising a plurality of sets of images and scene cuts. The invention also relates to an encoder, said encoder implementing said method.

Such a method may be used in, for example, a video communication system.

BACKGROUND OF THE INVENTION

A video communication system, like for example a television communication system, typically comprises an encoder, a transmission medium and a decoder. Such a system receives an input digital video signal, encodes said signal thanks to the encoder, transmits the encoded signal also called bit stream via the transmission medium, then decodes or reconstructs the transmitted signal thanks to the decoder resulting in an output digital video signal. A digital video signal comprises some sets of images and scene cuts.

Each image of the digital video signal is encoded along different schemes: either in an intraframe model, or in an interframe one as described in the standard MPEG2 referenced ISO/IEC 13818-2:1996(E), "Information technology – Generic coding of moving pictures and associated audio information: Video", International standard, 1996. In order to take into account the scene cuts, the encoder uses statistics codes, well known by the person skilled in the art, and encodes the images following a scene cut with reference to the statistics codes. At the decoding side, the decoder decodes the images. The scene cuts appear automatically thanks to the previous encoding.

One inconvenient of this encoding process is that, it is difficult to highly improve the rate/distortion ratio whatever the encoding scheme is used, wherein the rate/distortion ratio is the bits rate used for encoding versus the distortion perceived in the decoded image compared an original image.

SUMMARY OF THE INVENTION

Accordingly, it is an object of the invention to provide a method and an encoder for processing a digital video signal, said digital video signal comprising a plurality of sets of images and scene cuts, which allow an improvement of the ratio rate/distortion.

To this end, there is provided a method comprising the steps of:

- Localizing said scene cuts, and
- Issuing a set of images just after a scene cut by calculating said set from a visually distinguishable image after said scene cut.

In addition, there is provided an encoder comprising:

- Localization means for located said scene cuts, and
- Calculation means to issued a set of images just after a scene cut, said set being calculated from a visually distinguishable image after said scene cut.

As we will see in detail further on, the invention is based on the fact that under standard viewing conditions, human eyes cannot distinguish very fast changes in scenes. Therefore, with this principle, the encoding method according to the invention takes into account visually distinguishable images to encode the previous images after a scene cut. Therefore, only relevant part of information, the distinguishable images, is encoded as usual, whereas non-relevant-part of information, the non-distinguishable images, may be degraded or omitted. The rate/distortion ratio is then improved.

Advantageously, in a first non-limited embodiment, the calculation for the set of images is done by applying on said set a same encoding value than for said visually distinguishable image. In this embodiment, the calculation is very easy and very fast and does not need a complex system. The human eye can't see any difference.

Advantageously, in a second non-limited embodiment, the calculation for the set of images is a well approximation by general coarse motion compensation of said visually distinguishable image. In this embodiment, the calculation is very easy and very fast and gives a better result than the first embodiment, as there will have less fixed images and a well-informed human eye will be totally put on.

BRIEF DESCRIPTION OF THE DRAWINGS

Additional objects, features and advantages of the invention will become apparent upon reading the following detailed description and upon reference to the accompanying drawings in which:

- Fig. 1 illustrates a video communication system comprising an encoder according to the invention,
- Fig. 2 is schematic diagram of a first encoding of a digital video signal comprising images and a scene cut, applied by the encoder of Fig. 1, and
- Fig. 3 is schematic diagram of a second encoding of a digital video signal comprising images and a scene cut, applied by the encoder of Fig. 1.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, well-known functions or constructions by the person skilled in the art are not described in detail since they would obscure the invention in unnecessary detail.

The present invention relates to a method for processing a digital video signal, said digital video signal comprising a plurality of sets of images and scene cuts, said method is used in particular in an encoder ENC as shown in Fig.1 within a video communication system SYS. Said system receives some digital video signals.

In order to transmit efficiently some video signals through a transmission medium CH, said encoder ENC applies an encoding along different schemes well-known by the person skilled in the art: either in an intraframe model, or in an interframe one. Then the encoding signal known as bit stream is sent to a decoder DEC, which decodes said signal.

Said encoder ENC comprises:

- Localization means M1 for located said scene cuts CUT, and
- Calculation means M2 to issued a set of images just after a scene cut, said set being calculated from a visually distinguishable image after said scene cut CUT.

The encoding is done as following.

In a first step 1) there is a localization of scene cuts CUT, generally with statistics codes, for indicating the place of each scene cut within the video signal. In addition, a flag is used to indicate if the images after said scene cut have to be coded as usual, by a DCT coding for example, or has to be degraded or omitted, as it is detailed hereinafter.

From this localization, we can define two sets of images, a previous one, which is before a scene cut CUT and a next one, which is following said scene cut CUT.

In a second step 2), the sets of images just after the scene cut CUT is calculated from a visually distinguishable image after said scene cut.

This encoding method is taking into account human eyes capabilities. Indeed, perceptual studies as described in the documents "B. Girod, The information theoretical significance of spatial and temporal masking in video signals, Proc. SPIE/SPSE Conf. on Human Vision, Visual Processing and Digital Display, Los Angeles, CA, USA, pp. 178-187, January 1989", and "B. Girod, How important is masking for picture coding? Proc. International Picture Coding Symposium PCS '88, Torino, Italy, pp. 1.2.1-1.2.2, September 1988", have shown that under standard viewing conditions well known by the person skilled in the art, human eyes can not distinguish very fast changes in scenes: it is called the temporal masking effect. Therefore, the encoding is based on the idea that as human eyes can not distinguish image details in the fraction of second following a scene cut (human

eyes need to get used for at least $1/10^e$ of second), this biological property may be exploited in terms of video coding: during the accommodation of the eye, all pieces of information do not need to be present in the images.

The images, which cannot be perceived correctly by the human eye, are called non-relevant, whereas the other images are called relevant.

Therefore, in order to encode as much images as before with far less bits, the encoding according to the invention, encodes, with the general scheme code, for example DCT coding, only the relevant images whereas non-relevant images may be degraded or omitted. The visual quality remains the same. The non-relevant images are the one just following a scene cut, which cannot be perceived by the human eyes.

For example, as illustrated in Figure 1, if a scene cut CUT happens between a first image I(t0-1) and a second image I(t0), we can suppose an image complete details will only be distinguishable on the third image after the scene cut CUT, i.e. I(t0+2).

Hence, in a first non-limitative embodiment, the calculation for the non-relevant images is done by applying on said set of images a same encoding value than for said visually distinguishable image I(t0+2).

As shown in Fig. 2, the 2 previous images I(t0) and I(t0+1) after the scene cut CUT can be well approximated by the third image I(t0+2). In that case, the encoding is I(t0-1), I(t0+2), I(t0+2), I(t0+2), I(t0+3), I(t0+4), etc. This set of images can be very efficiently coded because of the identical successive images. Note that a simple flag may signal that the image is simply repeated from a previous image, said flag being inserted in the bit stream. Thus, in the previous example, the image I(t0-1) will be coded, then I(t0+2), then there will be 2 copy flags and then the image I(t0+3) will be coded.

Another alternative is to have a simple flag that may signal that the image is simply repeated from a following image.

In a second non-limitative embodiment, the calculation for the non-relevant images is a well approximation by general coarse motion compensation of said visually distinguishable image I(t0+2), for example by mesh method well known by the person skilled in the art.

Thus, as shown in Fig. 3, the previous images I(t0) and I(t0+1) just after said scene CUT can be well approximated by general coarse motion compensation of the third image I(t0+2). In that case, the encoding is I(t0-1), I(t0+2)-d0, I(t0+2)-d1, I(t0+2), I(t0+3), I(t0+4), etc, with d0, d1 representing a coarse moving of the pixels between two images. This set of images can be very efficiently coded because of the general coarse motion

vectors used for motion compensation after the scene cut CUT. Note that a simple flag may signal that the image is simply approximated from a previous or following image.

Practically, in case of an image rate of 30 Hz, if human eyes need at least $1/10^{e}$ s to accommodate, it means that only the third image will be distinguishably seen. Therefore, the quality of the two images between the scene cut CUT and this time may be cleverly degraded as proposed above.

Note that, in case of slow motion within some sets of images in the video signal, the calculation of the non-relevant images, as described above in the two embodiments, can be applied on more than 2 images without shocking visual artifacts.

Thus, a first advantage of the present invention is to improve the rate/distortion ratio, without losing any perceptual quality, as the non-relevant information i.e. the non-distinguishable images are not encoded as usual, and so fewer bits are used.

The other the advantages of the present invention are, on one hand, to reduce the time taken by the encoding, as a copy or an approximation of an image is very fast, and on the other hand, to reduce the memory taken by the encoding process, and this without losing any perceptual quality (i.e. subjective quality) in the encoding.

It is to be understood that the present invention is not limited to the aforementioned embodiments and variations and modifications may be made without departing from the spirit and scope of the invention as defined in the appended claims. In the respect, the following closing remarks are made.

It is to be understood that the present invention is not limited to the aforementioned video application. It can be use within any application using a system for processing a digital video signal where the ultimate consumer being the human eye, such as applications including digital movies, HDTV, and transmission and visualization of scientific imagery. Image codes have to be designed to match the visual capabilities of the human observer.

It is to be understood that the method according to the present invention is not limited to the aforementioned implementation.

There are numerous ways of implementing functions of the method according to the invention by means of items of hardware or software, or both, provided that a single item of hardware or software can carries out several functions. It does not exclude that an assembly of items of hardware or software or both carry out a function, thus forming a single function without modifying the method of processing the video signal in accordance with the invention.

Said hardware or software items can be implemented in several manners, such as by means of wired electronic circuits or by means of an integrated circuit that is suitable programmed respectively. The integrated circuit can be contained in a computer or in an

encoder. In the second case, the encoder comprises localization means adapted to make the localization of a scene cut, and calculation means adapted to issue a set of images just after a scene cut, said set being calculated from a visually distinguishable image after said scene cut, as described previously, said means being hardware or software items as above stated.

The integrated circuit comprises a set of instructions. Thus, said set of instructions contained, for example, in a computer programming memory or in an encoder memory may cause the computer or the encoder to carry out the different steps of the decoding method.

The set of instructions may be loaded into the programming memory by reading a data carrier such as, for example, a disk. A service provider can also make the set of instructions available via a communication network such as, for example, the Internet.

Any reference sign in the following claims should not be construed as limiting the claim. It will be obvious that the use of the verb "to comprise" and its conjugations do not exclude the presence of any other steps or elements besides those defined in any claim. The article "a" or "an" preceding an element or step does not exclude the presence of a plurality of such elements or steps.

CLAIMS

1. A method for processing a digital video signal (VS), said digital video signal comprising a plurality of sets of images and scene cuts (CUT), characterized in that it comprises the steps of:
   - Localizing said scene cuts (CUT), and
   - Issuing a set of images just after a scene cut (CUT) by calculating said set from a visually distinguishable image after said scene cut (CUT).

2. A method for processing a digital video signal (VS) as claimed in claim 1, characterized in that the calculation is done by applying on said set of images a same encoding value than for said visually distinguishable image.

3. A method for processing a digital video signal (VS) as claimed in claim 1, characterized in that the calculation for said set of images is a well approximation by general coarse motion compensation of said visually distinguishable image.

4. A computer program product for an encoder (ENC), comprising a set of instructions, which, when loaded into said encoder (ENC), causes the encoder (ENC) to carry out the method claimed in claims 1 to 3.

5. A computer program product for a computer, comprising a set of instructions, which, when loaded into said computer, causes the computer to carry out the method claimed in claims 1 to 3.

6. An encoder (ENC) for processing a digital video signal (VS) with a relevant encoded method, said video signal comprising a plurality of sets of images and scene cuts (CUT), characterized in that it comprises:
   - Localization means (M1) for localize said scene cuts (CUT), and
   - Calculation means (M2) to issued a set of images just after a scene cut (CUT), said set being calculated from a visually distinguishable image after said scene cut (CUT).

7. An encoder (ENC) for processing a digital video signal (VS) as claimed in claim 6, characterized in that said second means issued a set of images just after a scene cut (CUT) so that said images have a same encoding value than said visually distinguishable image.

8. An encoder for processing a digital video signal (VS) as claimed in claim 6, characterized in that said second means issued a set of images just after a scene cut (CUT) so that said images are well approximated by general coarse motion compensation of said visually distinguishable image.

9. A video communication system, which is able to receive a digital video signal (VS), said signal being processed by the encoder defined in claim 6.

Method and encoder for processing a digital video signal

ABSTRACT

The present invention relates to a method and an encoder for processing a digital video signal, said digital video signal comprising a plurality of sets of images and scene cuts (CUT). It is characterized in that it comprises the steps of:

- Localizing said scene cuts (CUT), and
- Issuing a set of images just after a scene cut (CUT) by calculating said set from a visually distinguishable image after said scene cut (CUT).

Use: encoder in a video communication system

Reference: Fig. 2

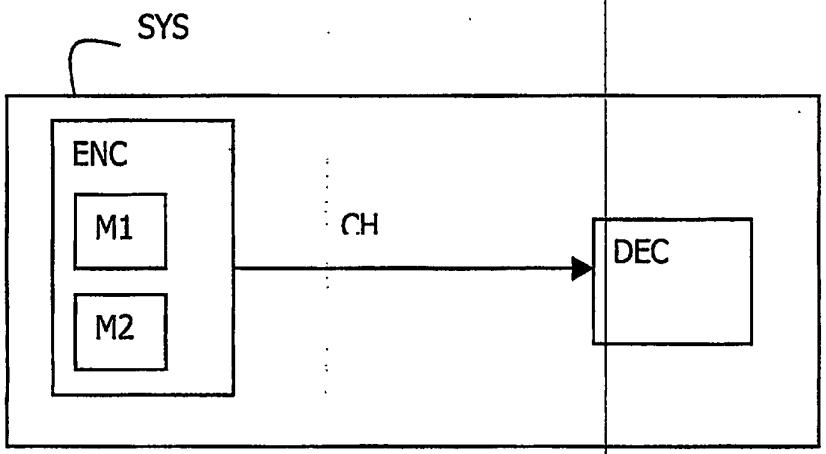

FIG. 1

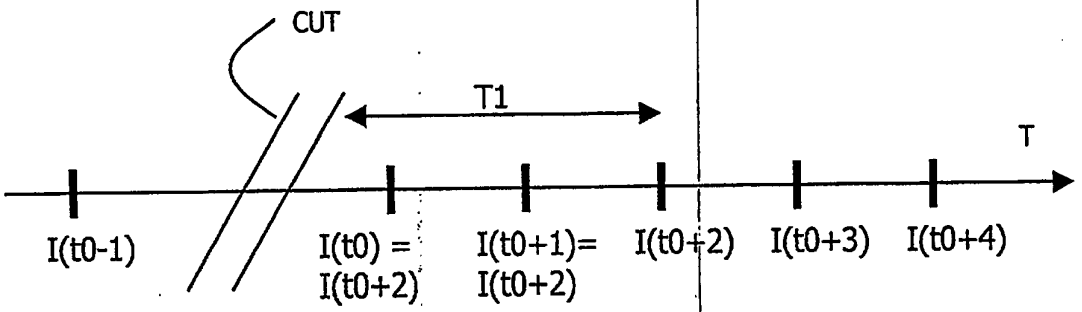
CUT
T1
T
I(t0-1)
I(t0) =
I(t0+2)
I(t0+1)=
I(t0+2)
I(t0+2)
I(t0+3)
I(t0+4)

FIG. 2

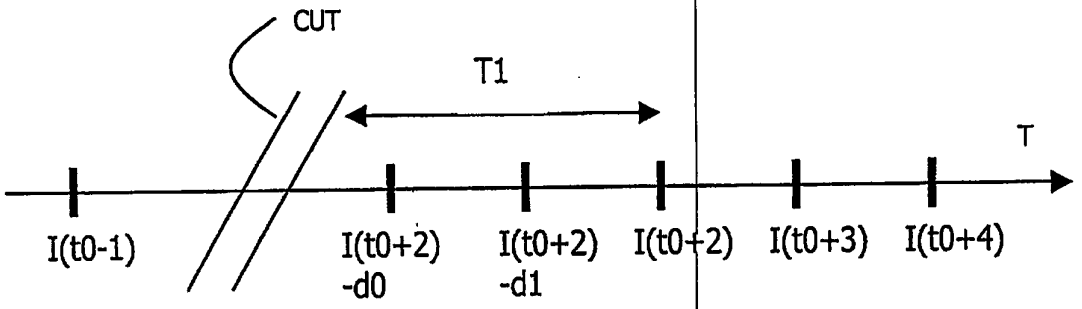
CUT
T1
T
I(t0-1)
I(t0+2)
-d0
I(t0+2)
-d1
I(t0+2)
I(t0+3)
I(t0+4)

FIG. 3

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.